OmniaScience

Intangible Capital

IC, 2019 – 15(1): 1-21 – Online ISSN: 1697-9818 – Print ISSN: 2014-3214

https://doi.org/10.3926/ic.1407

Trust under bounded rationality: Competence, value systems, unselfishness and the development of virtue

Natàlia Cugueró-Escofet[1], Josep M. Rosanas[2]

[1]*UOC (Spain)*

[2]*IESE Business School (Universidad de Navarra) (Spain)*

ncugueró@uoc.edu, jrosanas@iese.edu

Received January, 2019

Accepted February, 2019

Abstract

Purpose: This paper analyses the foundations of trust in a context of bounded rationality to reach the conclusion that non-calculative trust is meaningful essentially because of bounded rationality, specifying what aspects of bounded rationality are relevant for this to happen.

Design/methodology: Building on previous theoretical work we conceptually develop the reasoning involved to arrive deductively that bounded rationality provides a rationale for the concept of trust that goes beyond a calculative notion.

Findings: We show that there are four reasons for trust to exist and that people assess probabilities to each in order to determine whether to trust a recipient, depending on each of the four. We also add to previous work and show how bounded rationality provides additional arguments to show how competence, value systems and unselfishness are necessary to underpin trust. We provide additional foundations to their three factors, focused on bounded rationality. We add the development of virtue as a crucial fourth aspect, which supports the argument that trust can be reinforced between people and developed through time.

Originality/value: The concept of trust has been analyzed empirically, but it lacks some theoretical foundations to show under which assumptions trust is a requirement that goes beyond mere calculations, and can be developed or not through time. We also introduce how the concept of virtue has a major role in trust development.

Keywords: Trust, Bounded rationality, Virtues development, Ethics

Jel Codes: A13, D03, D21, L22, M14, M21

To cite this article:

Cugueró-Escofet, N., & Rosanas, J.M. (2019). Trust under bounded rationality: Competence, value systems, unselfishness and the development of virtue. *Intangible Capital,* 15(1), 1-21. https://doi.org/10.3926/ic.1407

1. Introduction

Trust has become a major topic in management. The special issues that major journals have devoted to the subject, like the 2003 special issue of Organizational Science, and the two special issues of the Academy of Management Review, in 1998 and in 2009, are signs of this importance. In Organizational Science, McEvily, Perrone and Zaheer considered that "although research on trust in an organizational context has advanced considerably in recent years, the literature has yet to produce a set of generalizable propositions that inform our understanding of the organization and coordination of work" (McEvily, Perrone & Zaheer, 2003, pp. 1). It seems reasonable to state that the field of trust in personal relations and organizations is wide and some central questions remain unexplored. This paper is intended to answer an important question that recent research in the field has considered to be basic: "Why do people trust each other?" (Dietz, 2011).

Different approaches have been used to attempt to find an answer to this question. This paper focuses on the "rational" approach, with the intention of showing that bounded rationality is necessary for the word "trust" to be meaningful; and that the reasons for the existence of trust arise from specific bounded rationality aspects. James (2002) showed what he called "the trust paradox", by which if rewards are changed so that both players have an incentive to take the action which is Pareto-optimal, the need for trust disappears. Rosanas (2016) went one step further to show that unbounded rationality necessarily leads to a calculative notion of trust, i.e., that if human beings were unboundedly rational, then Williamson (1993) would be right in that trust would simply be another name for risk. What we add here to his approach is the analysis of the aspects of bounded rationality that are at the origin of a meaningful concept of trust.

We think that revisiting trust and examining the assumptions that make the concept necessary will reveal fundamental differences between trust and risk. People trusting other people obviously face some risk, because this trust can be either honored or betrayed; and, as we have mentioned, trust has even been identified with risk (Williamson, 1993). However, risk and trust differ, and we are going to show that trust becomes important only if it can be distinguished from other concepts with which it shares some commonalities, risk being the main one.

Our endeavor here is twofold. First, we are going to examine trust compared to risk and second, we are going to show the assumptions that make the existence of trust possible. We think that this is relevant and useful as a starting point because the assumptions on which trust is based, and its distinction from risk, have not been examined following a rigorous theoretical analysis. We think our analysis is helpful because it allows to study the rational foundation of trust and, thus, it is an alternative way to the current empirical research and theoretical models that many researchers have started, which has already led to a better understanding of trust creation and reparation processes, and of their outcomes and effects (Dirks & Ferrin, 2001; Ferrin & Dirks, 2003; McEvily et al., 2003; Tomlinson & Mayer, 2009; Williams, 2007).

There is an additional reason why we think that this is an appropriate time to pinpoint the assumptions on which trust is based, as they are related to bounded rationality. Today, more than ever before, in the context of a crisis of values that has been at the root of many recent scandals, trust is potentially more useful than ever as the possible glue that might hold organizations together and increase the probability of improving the rectitude of current managerial practices (Hosmer, 1995). In general, definitions must have an intention of integration and, as such, the basic assumptions on which they are based must be explicitly included in those definitions. Hosmer does this by defining trust as the "expectation by one person, group or other firm of ethically justifiable behavior – that is, morally correct decisions and actions based upon ethical principles of analysis – on the part of the other person, group, or firm in a joint endeavor or economic exchange" (Hosmer, 1995, pp. 399). Hosmer considers trust at all possible levels of analysis, but here we are going to concentrate our rational foundations at the individual level of analysis, namely, trust between two people.

Furthermore, trust can promote the reinforcement of moral virtues and, thus, increase the probability that companies may enter into a process of inter- and intra-organizational trust building. Dietz argues that institutionally based trust is not possible alone; trust sources can come from institutional and interactional sources, meaning that now, more than ever before, "cultures depend on a more values-based leadership where people don't need to look at the rule book, where they know intuitively what the right thing to do is" (Dietz,

2011, pp. 218). This is crucial to the development of virtues: the required self-control and self-empowerment that creates trustworthiness increases as a result of a greater commitment to virtues and to the personal integrity of leaders. Cultural differences cause people to view trust differently. Different corporate cultures, depending on the industry of operations, also reveal different approaches to trust (Ferrin & Gillespie, 2010). This makes reaching an agreement on trust an aspect that organizations need to support. It also seems to be a transversal characteristic, even if trust formation can be reached through singularities that are dependent on cultural differences.

Some of the analyses of trust are based on the assumption that trust is rationally based and instrumental (Kramer & Tyler, 1996, pp. 10); but this instrumental view of trust is not enough to explain its presence in people relationships . In fact, and again according to Kramer and Tyler, "trust is important only when people have social relationships" (Kramer & Tyler, 1996, pp. 10). Besides, as we will see, people often adopt some type of rule-based decision-making, based on their own identities as individuals (March, 1987) and their identification with a group, possibly to protect existing social values and relationships, even in purely economic terms (Kahneman, Knetsch & Thaler, 1986).

As we mentioned, our starting point is the Rosanas (2016) result that unbounded rationality and a concept of trust that goes beyond risk are incompatible, i.e. that given unbounded rationality, trust is reduced to a mechanical form of calculation. In this paper, we intend to contribute to an integration of different concepts of trust and we do so formally developing sufficient conditions for trust to exist. We thus attempt to establish the decision-theoretical bases on which trust can be founded, showing that ill-known payoffs and preferences first, and values and value systems next, are essential for a concept of trust that goes beyond risk.

The paper is organized as follows. First, we briefly review the definitions of trust that can be found in the literature. Second, we investigate those situations in which trust between people seems to be a variable worth considering, showing that the analysis of bounded rationality provides a reason for a concept of trust that goes beyond the calculative notion discussed before. And we finish by arguing how value systems provide an underpinning for trust based on integrity, while having unselfish values provides an underpinning for trust based on those values and time-consistent preferences provide an underpinning for trust based on character or virtue. Finally, we relate the instrumental-rational approach to trust to the social and cultural approaches.

2. Defining trust and identifying the problem

The concept of trust has been considered "elusive" (Arrow, 1974; Gambetta, 1988; Williamson, 1993); so, finding a proper definition of trust is not an easy task (Gambetta, 1988; Williamson, 1993). It was originally defined as simply telling the truth or keeping one's promises (Dasgupta, 1988). Williamson considers that trust is just another word for risk taking and that, therefore, all trust is merely calculative (Williamson, 1993). He thus attaches no special importance to the concept, although later on in the same paper he accepts that there are other forms of trust (what he calls hyphenated trust and nearly non-calculative trust) that go beyond the mere calculative notion. On the other hand, trust has been investigated by behavioral researchers who have considered it to be linked to openness between two people and the extent to which one person can predict the behavior of others, especially with respect to what is normally expected of a person acting in good faith (Gabarro, 1978, pp. 294).

There is one definition that seems very useful for the purpose of the present paper: the one provided by Zand (1972). He defines trust as "actions that (a) increase one's vulnerability, (b) to another whose behavior is not under one's control, (c) in a situation in which the penalty (disutility) one suffers if the other abuses that vulnerability is greater than the benefit (utility) one gains if the other does not abuse this vulnerability" (Zand, 1972, pp. 230). Thus, the decision whether to trust comes first (when the trustor evaluates the trustworthiness of the trustee), and the decision whether or not to honor that trust comes afterwards. The essence of Zand's definition has been adopted by many authors, except perhaps for point (c), which has not been considered crucial to many aspects of the analysis. As we will see, Kreps (1990) adopted a similar concept when he formalized the analysis of trust as a "one-sided version of the prisoner's dilemma game" (Kreps, 1990, pp. 101).

Dietz believes "this basic sequence to be a universal dynamic, common to all trust encounters. (…) There is always an assessment (however thorough) of the other party's trustworthiness which informs a preparedness to be vulnerable that, in genuine acts of trust, leads to a risk-taking act" (Dietz, 2011, pp. 215). Some aspects may change, in contrast, in how the process is contextually configured. This mainly depends on how the evaluation of trustworthiness takes place and which variables are concluded to be more or less important in each part of the process. In general, trust has been considered a consequence of trustworthiness, i.e., mainly a perception of trust placed on a person who is evaluated as to whether he/she deserves to be trusted or not.

In general, most researchers recognize that the literature on trust relies heavily on the fact of being vulnerable or assuming some risk associated with the actions others take (Korczynski, 2000), and it seems that there is no doubt cast on the fact that trust is a risk-taking activity (Dietz, 2011). Korczynski stresses the point that the literature considers that there are multiple types of trust – such as rational calculative trust, or altruistic or blind trust – and that there is a distinction to be made between personal trust and trust in abstract systems and institutions (Korczynski, 2000, pp. 3). Recently, as we mentioned before, researchers like Dietz have considered that it is not the case that many types of trust exist. They have claimed that there is a universal trust experience or process, and what differs is how each stage of this universal trust sequence happens (i.e. considering different cultural or personal characteristics, among many others). The differences in how this process performs is apparent in different evaluations of trustworthiness, cognitions and actions of trust, and will thus originate different effects coming from a trust experience (Dietz, 2011).

Rousseau et al. in an article reviewing trust, have considered how researchers agree or disagree with the meaning of trust. (Rousseau, Sitkin, Burt & Camerer, 1998). They consider that the specific conditions for trust to exist are two: 1) the existence of risk and 2) interdependence, which leads to a definition of trust being "the willingness to be vulnerable in conditions of risk and interdependence" (Rousseau et al., 1998, pp. 395). And they continue by stating that "trust is not a behavior (e.g. cooperation) or a choice (e.g. taking a risk), but an underlying psychological condition that can cause or result from such actions" (Rousseau et al., 1998, pp. 395). Our focus is to discuss this in terms of rational foundations for trust, so discussing when trust becomes necessary or not. Even in a situation of risk and interdependence trust may not be necessary if mere calculativeness is enough (Rosanas, 2016). We discuss here that there is a third aspect joining risk and interdependence: bounded rationality, whose different aspects are going to discuss here to complement the previous two foundations of trust.

2.1. Formal modeling of trust

Kreps (1990) provided what might be the best formalization of the trust process in the form of a game. He based his formulation on the following points. First, the situation involves two people: the trustor (labeled A), and the trustee (labeled B). There is a sequential decision-making process, where A first makes an evaluation about B's trustworthiness; after that, A may be willing to trust B. If A makes the decision to trust B, then B can make the decision whether to honor that trust or betray A. The payoffs to both A and B will be determined by this decision. Of course, if A does not trust B, B can do nothing (see Figure 1).

"Vulnerability" in the Kreps model, means that if B betrays A, A's payoff is negative, i.e., A is worse off after the interaction than he/she was before. In contrast, if B honors A's trust, then A's payoff is positive. As we mentioned before, it is important that the payoffs to B are such that they are positive under both situations (B betrays and B honors), but are higher if B betrays. Otherwise, the only problem would be one of professional competence, with no possible conflict between the two agents. In that case, there would be no need for personal trust.

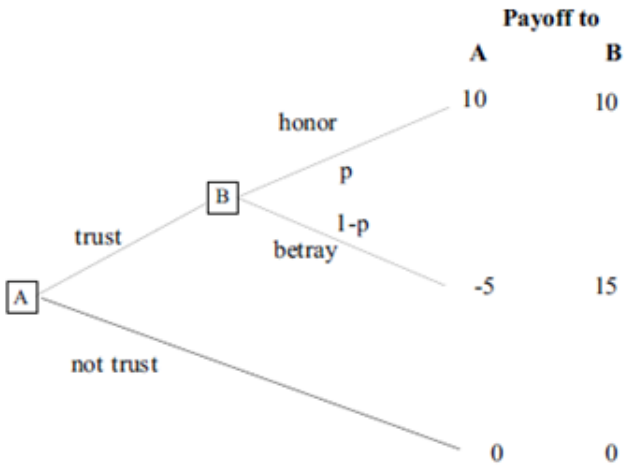

Figure 1. A quantitative representation of the trust problem

2.2.Trust and trustworthiness

The most widely used definition of trustworthiness is the one that was established in an integrative paper by Mayer, Davis and Schoorman (1995). They consider that people mainly evaluate three aspects of the trustee in order to assess his or her trustworthiness: ability, benevolence and integrity. They consider the trustor's willingness to enter into a risk-taking activity to be a central aspect, as trust entails the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995, pp. 712). The Mayer et al. definition misses an important aspect of trust, though: in order for the situation to be meaningful, the potential trustee has to have something to gain by performing an action that is not favorable to the interests of the trustor. If not, the interests of the two people are perfectly aligned and thus, in general, there should be no problem. This notion is intended to overcome the notion of opportunism, which is defined by Williamson as "self-interest seeking with guile" (Williamson, 1985, pp. 30). A situation of trust implies the belief by the trustor that the behavior of the trustee is not going to be opportunistic.

In a more recent paper, Schoorman et al. revisited and commented on the widely examined model of trustworthiness and trust they developed in 1995 (Schoorman, Mayer & Davis, 2007). They considered which new aspects should be included to advance the research on trust. Specifically, they proposed that some new trust topics that need to be developed include: time frame models, the relationships between trust and control systems as managers of risk, including variables that can show context dependence, setting up the role of trust reparation policies after a violation of trust, deciding how attachments and emotion can have an impact on trust, and finally studying how different cultures differ in their assessments of the ability, benevolence and integrity of trustees (Schoorman et al., 2007, pp. 352). They have not changed their model in Mayer et al., and have argued that even if trust is a multilevel experience, they value their model because it is simple, yet it includes many of the stages that the trust process entails, especially regarding the three dimensions of trustworthiness, which remained unchanged.

These three dimensions of trustworthiness proposed by Mayer and colleagues (ability, benevolence and integrity) generate what the authors define as trust, moderated by the trustor's propensity to trust (Mayer et al., 1995, pp. 715). The trustor evaluates these three aspects with respect to the trustee on the basis of a specific situation. The authors draw a parallel between these three characteristics and the foundation of convincing in Aristotle's Rhetoric, where he suggests that a speaker's ethos (the speaker's power of evincing a personal character which will make his speech credible) is based on the listener's perception of intelligence, character and goodwill. Ability corresponds to intelligence, but it is domain-specific; character (reliability, honesty) corresponds to integrity; and goodwill corresponds to benevolence, which is the perception of the trustee's positive orientation toward the trustor.

It is important to stress that while the behavioral literature puts the emphasis on how trust is detected through perceptions, we are more interested in showing how the reasons for trust are present and useful in practical situations (i.e., whether a situation is centered on trust, how to face risk and how to enable cooperation, among many other possibilities). The presence of trust involves some assumptions. Our paper deals with trust foundations and how this connects with concepts of trustworthiness perceptions, which may come afterwards, and concepts that come after trustworthiness, such as risk-taking activities (see Figure 1 in Mayer et al., 1995, pp. 715).

Our approach is consistent with the literature that considers trust as a multilevel relational process that needs more guidance and specification (McEvily, 2011), and we support the thesis of that article: that trust maters. The antithesis, that trust does not matter, is true only in the case of unbounded rationality, with its unlimited capacity for calculation and self-knowledge.

We thus try to focus on making trust assumptions explicit, at the same time distinguishing trust from the concept of risk by making trust broader than mere calculative trust, and explicitly showing the distinction between the two types of trust.

In order to do that, we next turn to examining the basic framework for the trust experience. We focus on a dyad of trustor and trustee, even if we are willing to accept that people's assessment of the trustworthiness of another individual is influenced by aspects that also come from the organization in which the relationship occurs. The dyad approach is a simplification which is valid for our purposes of clarifying trust assumptions and defining trust more clearly, since it is not situation specific, but can be extended to different types of relationships that are representative enough with the dyad as a symbolic relationship.

2.3.Bounded rationality: Ill-known payoffs and preferences

Most analyses of bounded rationality focus on the limited ability of human beings to derive (or calculate) the consequences of their actions. Some formalizations attempt to go further (e.g. by including a level of aspiration that is satisfying instead of attempting to maximize utility, see Selten, 1999). Herbert Simon's original formulation of the concept was wider, and included two additional characteristics of human thought. Aside from human beings' limited ability to foresee the consequences of their actions and the logical implications of their thoughts, they also have a limited ability to anticipate whether they will like the consequences of their actions, or how much they will enjoy the actions themselves: "It is a commonplace experience that an anticipated pleasure may be a very different sort of thing from a realized pleasure" (Simon, 1997, pp. 95). Finally, human beings have a limited ability to find possible courses of action as solutions to problems, i.e., the alternatives for action are not out there waiting for someone to pick them up; they have to be generated, often at a considerable cost and/or effort.

In order to analyze the role bounded rationality plays in the problem of trust, we need to focus on the first two of these limitations. The first one essentially means that, in complex situations, agents may not be able to know the actual payoffs to themselves, not only because of external, or objective uncertainty (which may be present, of course), but because they are not able to figure them out, or determine their probabilities with any degree of accuracy. One of the characteristics of unbounded rationality involves being able to accurately assess the probabilities of uncertain external events. Contrary to an extended belief, uncertainty is perfectly compatible with unbounded rationality provided that the agents can make an accurate evaluation of the probability of uncertain events and are able to evaluate the results of their actions with (again) an accurate estimation of the probability of those results obtaining (and also, as we show next, the utility they derive from those results given the risk). The inability to assess such probabilities includes the possibility of being unable to foresee certain specific consequences; or, in more formal terms, the possibility that agents may assign a zero probability to events that are perfectly conceivable, because they have simply overlooked them. For A, this limitation represents uncertainties with respect to B's behavior in addition to the usual uncertainties about objective events.

The second characteristic is like an extension of the first one: agents are unable to accurately predict the subjective part of the payoff, i.e., the utility to themselves of the expected results. In other words, decision-

makers have to anticipate future preferences of which they cannot be sure. March states this distinction very clearly: "The conception of choice enshrined in the axioms of contemporary decision theory and microeconomics assumes optimization over alternatives on the basis of two guesses. The first guess is about the uncertain future consequences that will follow from alternative actions that might be taken. The second guess is about the uncertain future preferences the decision-maker will have with respect to those consequences when they are realized" (March, 1987, pp. 155).

In the context of trust, the implications go beyond the uncertainty added to A about B's behavior, which has already been considered above. The problem is particularly interesting when the explicit, quantifiable payoffs are of the kind that lead to the trust problem. Formal examples of this can be seen in Rosanas (2016). In these cases, the trustee has an explicit incentive to betray. But what value do the non-quantitative variables have for the trustee? Under unbounded rationality, the answer is quite clear (at least to him/her). Thus, if the non-quantitative variables can be perfectly valued, as they are in Rosanas (2016), the problem ends there, and Williamson (1993) is right. In contrast, with bounded rationality, it may well be that honoring A's trust is in B's best interest, but B him/herself may not know it (or, at least, not without some doubt or fuzziness).

Also, under bounded rationality, B may be sorry after making his/her choice, whatever that choice may be, which may change the alternative B chooses next time, if there is a next time. For instance, when making his or her initial decision, B may magnify the importance of the quantitative variables and betray A, only to realize – too late – that the non-quantitative variables were at least as important as the quantitative ones. Of course, the opposite may happen as well: on reflection, B may decide to honor A's trust because of the qualitative variables, only to find later on that they were not that important after all.

Human beings may face the same problem over and over again through time and make different decisions because of learning, or because of different states of mind (emotion, for instance) that bring some aspects of their lives and some of their values into the focus of their attention to the relative neglect of others (Loewenstein, 1996; Simon, 1983; Simon, 1987). In the presence of a time constraint (which would be irrelevant in the case of unbounded rationality), when decision time is scarce, an optimization approach may not be feasible for an agent who is not so familiar with the problem (Selten, 1999).

The ability to optimize may not be symmetrically distributed in a problem involving trust (i.e., one of the agents may be more familiar with the problem and how to solve it than the other). To be specific, if A is more familiar with the problem than B, A may think it is a possibility that B will choose wrongly according to B's own preferences, harming A in turn. The possibility of A trusting B is obviously affected by this type of assessment. Persuasion may play an important role in those situations.

Hirschman (1984) has argued that there are two kinds of activities. Some human activities are instrumental, and are done in order to get a paycheck or an explicit reward. Others are not: those activities that are undertaken for their own sake or that carry their own reward fall into this category. Some activities have such an uncertain reward that they will seldom be undertaken for that reason. There is an education process in such activities, however. Not everybody likes them, only those people that have 'learned to love them'. This is unlikely to happen with simple, routine types of jobs or activities. But it is much more likely in more complex situations, precisely where trust is of higher importance than in rather simple contexts. The Hirschman analysis includes other factors, like the willingness to put plans of action into practice, which will be considered below.

This is of course related to the classical distinction in the behavioral sciences between "intrinsic" and "extrinsic" motives. This distinction comes from the literature of the 1950s and 1960s (see, e.g., Lawler, 1969; Saleh & Hyde, 1969). Ryan and Deci (2000) and Lindenberg (2001) distinguish between "intrinsic motivation, which refers to doing something because it is inherently interesting or enjoyable, and extrinsic motivation, which refers to doing something because it leads to a separable outcome" (Ryan & Deci, 2000). Frey (1997) and Osterloh and Frey (2003), consider that intrinsic motivation may have a hedonic component of enjoyment, while at the same time there is a normative intrinsic motivation out of a sense of obligation. We add to that the possibility of learning, i.e., that the intrinsic motivation may not be apparent from the beginning, but may have to be learned or acquired (as we will see below) by persuasion.

In general, preferences may change through time, depending on each agent's experience. An agent may 'learn to love' some variables or situations, and 'learn to hate' others. Intangible variables found in business situations, like the value placed on personnel development, the public image of the company, or the internal human climate of the firm, may change dramatically through time, as agents learn about their jobs, about the organization and about themselves.

An important aspect of this learning process is persuasion. In a world of unbounded rationality, there is no place for persuasion: every agent knows his/her preferences perfectly (including the evolution of preferences through time) and there is no reason to change. By contrast, in a world of bounded rationality, while some variables (monetary rewards, for instance) are easy for everybody to appreciate, persuasion may play an important role in 'learning to love' some goals or activities. Barnard (1938) already stressed the importance of persuasion as one of the crucial methods (together with incentives) to get people to work in the interest of the organization. At present, common everyday experience, as well the momentum of communication courses in business schools, can be said to confirm Barnard's intuition.

In the context of the analysis of trust, the importance of the imperfect knowledge of one's own preferences, which implies the kind of learning just mentioned, cannot be overstated. But with bounded rationality, even if both the monetary and non-monetary results are well-known (perhaps with some uncertainty), the utility that agents derive from them is not well-known, and there may be some need for persuasion in order to get an employee (or an external stakeholder) to cooperate. If we go back to Figure 1 and we consider that, aside from the explicit results shown, there are also some non-monetary results, B may be indifferent between the results derived from betraying or honoring trust; however, B may be persuaded (by A or by anyone else) that he/she will be better off with the results when honoring the trust. If A believes B has been persuaded, A may trust B to achieve a result that is optimal economically and otherwise. But, of course, if B is disappointed, his/her attitude next time around will surely be different. So, persuasion needs some element of truth to work in repeated interactions.

2.4. Systems of preferences and values

According to Simon, bounded rationality is the type of behavior that is "intendedly rational, but only limitedly so" (Simon, 1957, xxiv). Typically, it is intendedly rational through a system of preferences and values. In economic theory, the preferences of individuals are considered to be "arbitrary". In economic theory, rationality essentially means the consistency of those preferences (in technical terms, transitivity of preferences) to avoid circularities; but, except for that aspect, the preferences of two individuals between, say, two goods, may be completely unrelated. In a world of unbounded rationality, a preference map is a perfect reflection of the individual's preferences and values, and there is no distinction between different levels of values by importance or by familiarity with the specific situation.

Under bounded rationality, however, values may look substantially different. Herbert Simon borrowed the distinction between fact and value from the logical positivistic philosophy (Simon, 1997, chapters 1, 3 and 4). In accordance with that philosophy, he related rationality to the choice of means conducive to the achievement of previously selected goals (Simon, 1997, pp. 4). The selection of goals itself would not be rational or irrational, then, it would just be a matter of taste on the part of the individuals.

Simon recognized, though, that there is a "hierarchy of decisions", a means-ends chain so that a given goal is often a means to achieve a higher end: "Ends themselves are often merely instrumental to more final objectives (…) Rationality, then, has to do with the construction of means-ends chains of this kind" (Simon, 1997, pp. 73). Therefore, whether a final end is achieved through an intermediate end or not is entirely a matter of fact, susceptible of being tested empirically. For many human beings, most of the means pursued actively are only means to higher ends.

Preferences, then, operate essentially on the higher ends; and the lower ends are obtained as a consequence. Rokeach (1973) claims that people have relatively few basic values. Then, according to Fischhoff, Slovic and Liechtenstein: "If, as Rokeach claims, people have relatively few basic values, producing an answer to a specific

value question is largely an exercise in inference. We must decide which of our values are relevant to the situation, how they are to be interpreted, and what weight each is to be given" (Fischhoff, Slovic & Lichtenstein, 1988, pp. 401).

In particular, this applies to the context of the results to be achieved in a business firm, or in any organization in general. Profitability (or financial equilibrium) is always one of the crucial variables desired by managers, but at the same time they also value market position, or competitive advantage, or organizational knowledge, perhaps as a means to achieve the higher end of profitability, or perhaps even as higher ends in and of themselves. It is often claimed that, in business firms, profitability is the overriding goal and that all other ends have to be considered "intermediate" and evaluated instrumentally as means. However, other goals may also rank high in the hierarchy, such as social responsibility or personnel development.

Individuals, however, may not be able, or even willing, to make decisions in accordance with the relatively few basic values. Bounded rationality limits their ability to do so, and they have spontaneous impulses that may pull them in a different direction from those few basic values. Hence, their ability to make choices that are logically consistent with them is also limited. However, these limitations differ depending on the situation: in familiar situations, individuals are better able to optimize with respect to higher ends, and, thus, they may have very definite preferences in relation to lower ends: "People are more likely to have clear preferences regarding issues that are familiar, simple and directly experienced. Each of these properties is associated with opportunities for trial-and-error learning, particularly such learning as may be summarized in readily applicable rules or homilies. Those rules provide stereotypic, readily justifiable responses to future questions of values. When adopted by individuals, they may be seen as habits; when adopted by groups, they constitute traditions" (Fischhoff et al., 1988, pp. 399).

This may be seen as a specific instance of Simon's view about the role of intuition and emotion. Simon considers intuition as a shortcut in the chain of reasoning, in situations that are familiar, or that ring a bell with respect to similar experiences from the past (Simon, 1987). The Fischhoff et al. (1988) quote above may thus be seen as the application of that approach to the selection of values: for familiar situations, individuals know (or think they know) what preferences are coherent with the higher values; but in more unfamiliar situations, they might not. Fischhoff et al. (1988) also provide an interesting list of the states of mind associated with not knowing what you want in less familiar situations and some of the actions that follow (Fischhoff et al., 1988, pp. 400). That list is adapted by the authors in the form of a decision tree in the Appendix of this document.

We can see that having a coherent opinion and accessing it properly (the implicit assumption in fully rational models of behavior) is only one possibility among many. Knowing what you want is, one might say, almost the anomaly. It is perfectly possible for someone to have no opinion, or an incoherent one, not to realize it, and make decisions is spite of that. Living with incoherence is another possibility: in spite of higher ends being incompatible with some lower ends, people may try to achieve both (and fail, of course).

Applied to a situation involving trust between two people, the exercise in inference suggested by Fischhoff et al. in the above quotation would essentially consist in evaluating the possible dimensions of the consequences of each alternative, and evaluating to what extent the higher values are served. But the action to be initiated following this analysis may be in contradiction with lower ends, or immediate desires, or impulses, and the individual may be willing to be inconsistent. For instance, an individual may evaluate that Alternative 1 is better than Alternative 2 in terms of the higher values, but not in terms of the immediate monetary rewards. If those of Alternative 2 are bigger, the individual may be willing to live with the inconsistency.

In the next section we will go into the problem of willpower, by which one individual may really want to achieve a high objective, but be incapable of taking the corresponding action in the short run. For the time being, it is important to notice that we are not analyzing that problem yet, but only the problem at the cognitive level.

It follows from the list in the Appendix that rationality is not equally bounded for everybody. That is, some individuals try harder to be rational than others: some people are more reflective, and willing to check for coherence and act coherently with the higher ends. Others are more impulsive, or willing to pursue immediate,

lower ends without much reflection or need for coherence. For any specific individual, the probability that he/she is going to be more coherent is higher in familiar issues than in issues that are not as familiar, as stated in the above reference by Selten (1999) and the quote by Fischoff et al. (1988).

In summary, some people are willing to act coherently with their stable, higher values in spite of their short-term urges to do otherwise, and some are less willing. The type of behavior that consists of trying to be coherent with some stable, or permanent, set of higher ends is what is usually called integrity.

3. Different bases for trust

3.1. Trust based on personal competence

Personal competence is similar to what the behavioral literature on trust labels ability. It is a basic aspect of trust that has been defined by Bidault and Jarillo using the example of a surgeon: "Having trust in a surgeon goes beyond (only) thinking that he is honest. It is, more than anything else, believing that he will do his job well. The term 'well' in this context refers to patients' concept of 'state-of-the-art'" (Bidault & Jarillo, 1997, pp. 85). They continue: "The other party is expected to have the necessary skills to carry out the tasks specific to the transaction agreed upon" (Bidault & Jarillo, 1997, pp.85). They label this type of trust, "technical trust" and for them this includes the technical expertise associated with that profession together with the general competence that the profession requires. This concept is similar to what Mayer, Davis and Schoorman labeled "ability" (1995p.717). Following their definition, ability is "that group of skills, competences and characteristics that enable a party to have influence within some specific domain. The domain of the ability is specific because the trustee may be highly competent in some technical area, affording that person trust on tasks related to that area" (1995p. 717). Rosanas (2016) attempts to formalize that professional competence as the trustee having better information about the underlying phenomenon than the trustor. Technical expertise, ability or better information associated with professional competence are necessary conditions for A to be able to trust B in any circumstance: if B does not have that, he/she cannot be trusted to produce an outcome that is good for both sides. We summarize our analysis in Proposition 1:

Proposition 1: An essential element of trust is the professional competence of the trustee: technical expertise, abilities and better information are necessary for the trustee to obtain the desired results.

3.2. Trust based on integrity: Value systems

Of course, the necessary condition in our Proposition 1 is not sufficient. In a way that is consistent with Mayer et al. (1995), we will now review the role played by integrity.

According to Webster's Seventh New Collegiate Dictionary, integrity is precisely the "adherence to a code of moral, artistic or other values" (Webster, 1972, pp. 439). An important part of an individual's trust in other individuals will then reside in their integrity. We now turn our attention to this type of trust.

If we examine the trust problem in the light of the previous section on values and preferences, we see how trust becomes meaningful in our context of bounded rationality. The trustor, A, can make an (uncertain) assessment of the basic values of the trustee, B, and his/her willingness to make decisions as consistent with that set of values. That is what makes this situation different from the case of unbounded rationality: with unbounded rationality, any individual can see his/her whole preference map and all of the decisions he/she makes are by necessity consistent with that preference map. With bounded rationality, in contrast, there is an additional uncertainty about B's integrity, i.e., B's capacity to make decisions consistent with that basic set of values and preferences.

It is important to note that, so far, no assumptions have been made as to the content of the basic values. Thus, those values may be shared by A, or not. A may trust that B will (or will not) do something because of the basic set of values A assumes B has, not because A agrees with those values. Suppose, for instance, that A and B have different religious beliefs. A may know that B's religion forbids some practices, and that B is a strong believer. Then, A may trust B even if that action, forbidden by B's religion, is not considered bad by A at all. The example

of Yudhishthira in the Mahabharata epic, cited by Dasgupta (1988), is very much to the point here. Renowned for his trustworthiness, Yudhishthira once lied in order to throw off his unrighteous enemies. The lie worked because the enemies (who had a completely different value system that did not place value on trustworthiness) believed that Yudhishthira would not lie.

In the means-ends chains that human beings construct under bounded rationality, intermediate values are a means to more final values. Obviously, though, those higher values need to be consistent with each other; otherwise, a contradiction would require sacrificing one of them. In specific contexts, contradictions may arise. In this context, the Mahabharata example is particularly useful, because Yudhishthira would not lie to obtain immediate benefits for himself. Truthfulness is an important (higher) value for him, but he lies to defeat unrighteous enemies. Thus, according to Dasgupta, he qualifies as a consequentialist. A slightly different way of looking at the same problem is as a conflict between two values, both of which are rather high in the hero's preferences: truthfulness and the good of his people. When they are perceived as being incompatible, he resolves in favor of defeating the enemies, even if it is through a lie, because the good of his people is considered to be higher.

The basic set of values may be more or less volatile, depending on specific individuals and on their circumstances. But A's beliefs about B's set of basic values (the content of those values, how stable they are, and to what extent B is willing to make decisions based on them) will determine A's subjective probability p that B will honor or betray his/her trust. Then, in this context, trusting B implies A's belief that B is willing to make a given decision according to his/her entire set of values, beyond the explicit, monetary values involved in the decision. This is the concept of trust based on integrity, which can be expressed in the following proposition:

Proposition 2: An essential element of trust is the trustee's willingness to act in accordance with his/her system of values and preferences, because this makes the behavior of the trustee more predictable in specific situations.

3.3. Trust based on unselfish values

As we have seen, no matter what the values are and whether the trustor agrees with them or not, a person may be willing to make decisions that are consistent with them, or, on the contrary, a person may be rather impulsive. But, then, if we don't say anything else about the values, the result in terms of trust depends on the specific situation and the specific values involved. In other words, we cannot say that A trusts B in general, but A trusts B only under some specific circumstances and in some specific issue. Of course, the fact that the behavior of the trustee is predictable does not necessarily mean that the trustee will generally make decisions that are favorable to the trustor. For this to be true, we need to go one step further, to consider the kind of values that trustor and trustee have.

From this point of view, values can be purely selfish or altruistic; they may rate truth-telling, or fairness, or friendship, or social welfare, or the common good (the summum bonum of the Schoolmen) high or low. If B's values are selfish, then A can perhaps trust B about a specific problem, or for a small class of decisions only. In contrast, if B's values are non-selfish, A may be willing to trust B for a wide class of decisions.

Proposition 3: An essential basis of trust between two people under a variety of circumstances is that the trustee must have a system of values and preferences in which some non-selfish, or social, values (or, simply, the interests of other people) are placed high, and the trustee must be willing to make decisions according to such a system of values and preferences.

3.4. Trust based on the moral virtues of the trustee

In economic models of decision-making and organization (with unbounded rationality), it is typically assumed that people are impatient, i.e., that they like to experience rewards soon and costs later. This is captured in such models through the use of a utility function discounting utility over time exponentially. Such preferences are called time-consistent. But, in O'Donoghue and Rabin's words: "Casual observation, introspection, and psychological research all suggest that the assumption of time-consistency is importantly wrong. It ignores the

human tendency to grab immediate rewards and to avoid immediate costs in a way that our 'long-run selves' do not appreciate" (O'Donoghue & Rabin, 1999, pp. 103).

The problem of the discrepancy between a person's preferences at different times is also an old one in philosophy. The basis of Aristotle's criticism of Socratic ethics was that very point. In Aristotle's analysis, Socrates had said that "nobody acts in opposition to what is best if he has a clear idea of what he is doing. He can only go wrong out of ignorance." The reasoning, however, and again according to Aristotle, "is in glaring contrast with notorious facts": people may know what is right and not do it because of weakness of will, or lack of control (Aristotle, 2000, Book VII).

Schelling (1978; 1984) has studied the problem of time-inconsistency in depth. Wanting to quit smoking but not doing it; Christmas accounts that protect your money from yourself; free loans from the taxpayer to the IRS by understating the number of dependents; placing the alarm clock across the room; setting the watch a few minutes ahead to deceive oneself…"

"In these examples, everybody behaves like two people, one who wants clean lungs and long life and another one who adores tobacco, or one who wants a lean body and another who wants dessert. The two are in a continual contest for control: the 'straight' one often in command most of the time, but the wayward one needing only to get occasional control to spoil the other's best laid plan" (Schelling, 1978, pp. 290).

The two selves are not equally important in Schelling's analysis. He is obviously partial to the "straight" self. The section titled "Strategy and Tactic," for instance, in the 1984 paper, consists of recommendations so that the "straight" self can be in command of the "wayward" self: relinquish authority, let somebody else hold your car keys, order your lunch in advance, don't keep liquor or tobacco in the house, order a hotel room without a television, do your food shopping after breakfast...

Bazerman, Tenbrunsel and Wade-Benzoni (1998) formulate the problem in a slightly different way. They suggest that the two-selves theory can be conveniently made easier, into a "want/should" explanation, based on the empirical evidence available. When people are asked what they want, their responses will be emotional, affective, impulsive, and hot-headed; whereas when they are asked what they should do, their responses will be rational, cognitive, thoughtful and cool-headed. These are then the two selves: the "want self," and the "should" self.

In an approach that is complementary to the previous ones, to some extent, Loewenstein (1996) attributes the fact that people often act against their self-interest, in full knowledge that they are doing so, to "visceral factors" (hunger, pain, sexual desire, moods and emotions, etc.). Later, Kahneman examined in context the problems related with not knowing exactly what one wants (Kahneman, 2011).

In general, there is no doubt casted on the fact that human beings are sometimes incapable of doing what they think is in their best interest. Doing what one thinks one should do (what can be labeled "the dominion of the 'should' self over the 'want' self"), is what Aristotle called developing moral virtues or practical wisdom applied to oneself. The Aristotelian view looks at moral virtues being acquired with practice, what he called "practical wisdom." If that is true, one could expect that the impact of visceral factors, or the relative importance of the want and should selves, will depend very much on each individual and his/her past history regarding the personal development of moral virtue (or practical wisdom).

The organizational context makes things even more complex. A manager's self-interest may be substantially different from the (otherwise espoused) organizational objective. The manager may say, for instance, that the main goal of the firm lies in value maximization, while he or she may take actions that destroy long-term value at the same time, possibly for short-term benefit. Jensen (2000) and Senge (2000) provide excellent examples of that possibility. Jensen argues that this is the result of "the tendency of human beings to resort to short-term value-destroying actions in the name of value creation" (Jensen, 2000, pp.50). Indeed, and again according to Jensen, the latest financial scandals have only confirmed this tendency, even to an extreme degree (Jensen, 2002). This analysis leads naturally to the following proposition:

Proposition 4: A crucial element in one person trusting another is whether the second person is able to put into practice what he/she thinks will be best for him/herself in the long run, in spite of possibly attractive, short-term results.

The analysis in this section introduces a new facet of the word "trust." Previously, we consistently referred to the decision-making process as if all decisions made by an economic agent were to be immediately implemented with no problem. In the last section we suggested that this may not be so, and that the decision may change before implementation, not because of any new information coming in, or any changes in one's tastes, or any further thoughts on the basic values, but because of a lack of control of oneself. In terms of bounded rationality, it can be interpreted that the decision-maker's focus of attention shifts to the more immediate, attractive variables, in preference to future variables that are, in fact, preferable for an individual with unbounded rationality. To implement what one considers being the right or rational decision, willpower is needed. This is the Aristotelian point of view, cited in the previous section.

Different people at different points in time will have different degrees of willpower. According to Aristotle, this develops through practice. Ordinarily, an individual's willpower to do what he/she considers to be the right thing is too little (as in the Schelling examples), but Benabou and Tirole (2002) have shown how, under some conditions, people sometimes adopt excessively rigid rules that result in compulsive behaviors such as miserliness, workaholism, or anorexia. Quite obviously, this excess virtue is also acquired, as in the Aristotelian account, through practice. We summarize our focus in studying trust in Figure 2.

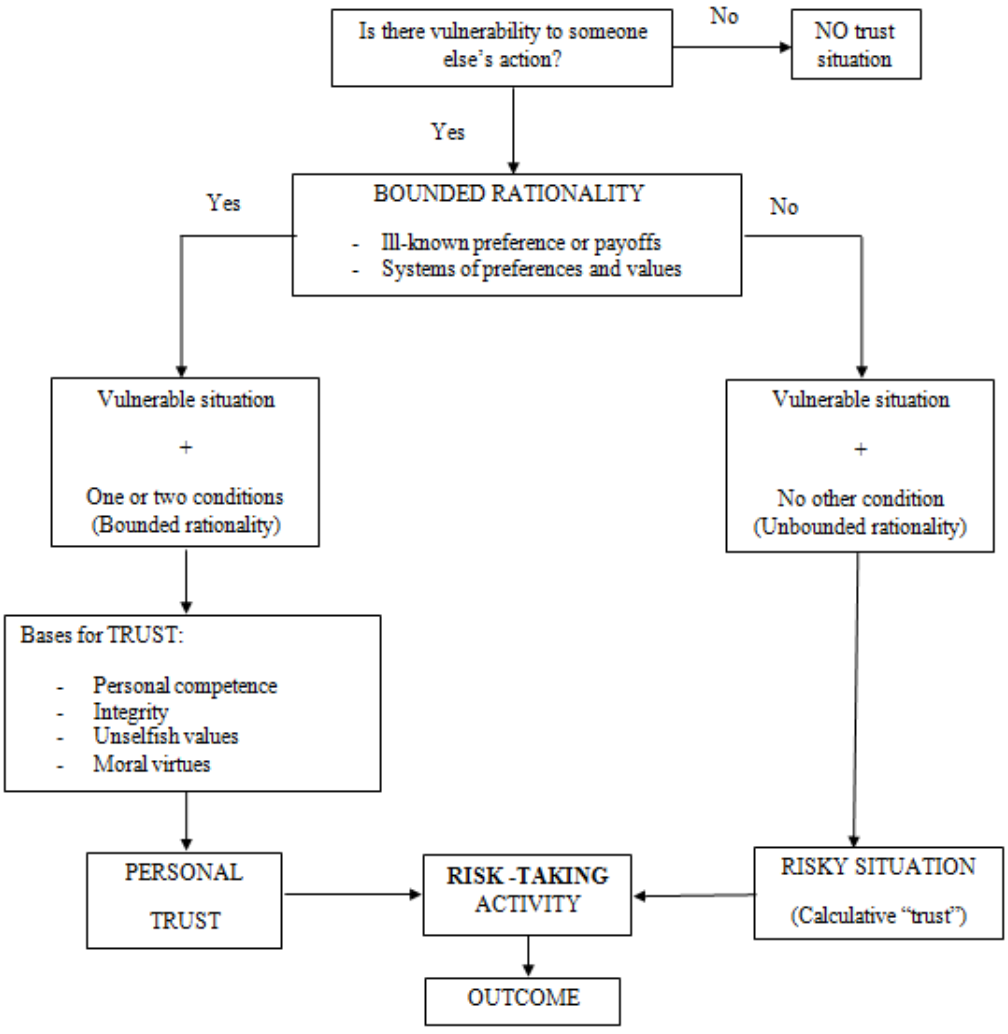

Figure 2. Assumptions for trust and antecedents of trust perceptions

3.5. The probability of B honoring A's trust

In fact, a trustor A, has to formally or informally assess the probability p that a trustee, B, will honor the trust. In fact, the trustor has to assess four probabilities: the probability p_c that the trustee has competence on the matter at hand, the probability p_v that the values of the other agent are stable and consistent, the probability p_s that such values are pro-social or at least not selfish and that B will choose to honor (possibly in spite of immediate material rewards for doing the opposite), and the probability p_w that B will in fact have the willpower to put that decision into practice, given that the decision has been made, i.e. (assuming they are independent):

$$p = p_c \cdot p_v \cdot p_s \cdot p_w$$

Obviously, p_v and p_s depend on B's system of values and on the willingness B has to make rational decisions according to that system; while p_w depends on B's willpower, and on the availability of attractive alternative actions to betray A. If there are no immediate, attractive rewards for B to betray A, so that B has no problem honoring A's trust, then p_w will be equal to 1; and the more attractive the rewards available to B for betraying A, the lower p_w will be. In summary:

Trust in another person is based on an assessment of four factors: (1) competence; (2) adherence of the other person to a (stable and consistent) system of values; (3) that such a system of values somehow includes social goals or the interests of other people; and (4) the willpower of the other person to put into practice what he/she believes he/she should. Let us stress, however, that an essential point in our analysis is that such an assessment cannot be done with any precision, for this would require unbounded rationality and we are in the opposite context. That is why trust cannot be merely calculative.

3.6. On the empirical evaluation of the p's

There is an important difference between the foundation of trust resting on judgment and competence and the foundation of trust resting on preferences and integrity. The former is entirely empirical: B cannot fake a knowledge that he/she does not have, and once he/she has it, he/she will continue to have it in the future. B, of course, may have to adapt to new situations in the future and learn more, and B may be luckier or unluckier in the short run, but the fact remains that the proof of B's competence is in the empirical success in the (average) results of his/her decisions.

The latter, in contrast, cannot be entirely empirical. It is empirical to the extent that all knowledge of the real world comes (obviously) from observation of empirical facts; but there is an important problem associated with assessing someone else's preferences and values: they can be faked. In repetitive decisions, one agent may fake a preference for some variables just to gain someone else's trust, then, once this is achieved, betray the other party. In fact, it is rather common for embezzlers to have an immaculate history of honesty, even to excess, until, one day, they betray the trust deposited in them (demonstrating, incidentally, that the trustor was actually vulnerable). So, for both reasons, even if a given person has shown unchanging preferences over a long period of time, one can never be sure that this behavior will continue indefinitely.

This is closely related to a well-known problem in philosophy, the problem of induction. Bertrand Russell (1959) remarked that the fact that, for ages, we have observed the Sun rising every day does not necessarily imply that it is going to rise tomorrow. The chances that the Sun will rise tomorrow vary greatly with the causal explanation we attribute to its motion. If the sun rises because some giants light a ball of fire at night every night and raise it in the morning, then, if one day they are too tired, or they feel whimsical, they may not light it at all. Yet, this was an explanation believed by some of our ancestors not too long ago, by historical standards. Currently, we believe in the laws of motion, gravitation and the Earth's rotation as causes of the sun rising, and this makes it much more unlikely that the sun will not rise tomorrow. Too many things would have to change. The will of the giants may change much more easily than the motion of enormous bodies and their laws.

Notice that the Russell example involves only physical systems (according to the explanation we accept today, one might add), where there are no reasons for doubting the regularity of the phenomenon. But if induction is a complex issue in the natural sciences, it is even more complex when the system under study is a human being.

The parable of the inductivist chicken provided (again) by Russell is very much to the point here: "Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken" (Russell, 1959, pp. 35).

In other words, mere repetition of a given choice by one individual is not a good basis to infer that it will be repeated again. Human beings are purposeful systems, and may have intentions completely different from the ones that seem obvious; and unless there is some understanding of the real reasons why they act, any forecast of their next action may turn out to be completely wrong.

Bounded rationality is again a determining concept. With unbounded rationality, human beings would have unlimited capacity for faking preferences; but they would also have unlimited capacity for accounting for that fact. That is, it would not be too difficult for one person to try to trick the others, but the others would immediately assign a probability to that eventuality and incorporate it into their subjective probability. The success of such a strategy would therefore be in doubt.

With bounded rationality, one agent may try to internalize the value and preference system of the other, partly through previous formal interactions of the same kind, but partly through other means of communication: words, body language, common friends, shared beliefs, attitudes… All these factors are relevant to determining the probability *p* that B will honor A's trust. The crucial fact, however, is that mere historical repetition of a given alternative in previous, similar situations ("reputation," if by that word we simply mean the other person's track record) is hardly enough. Internalizing the way the other person thinks and his/her value system is an important element in the assessment of the probability of the other person honoring or betraying trust.

3.7. Individualistic-rationalistic approach versus social and cultural approaches

This paper has taken the individualistic-rationalistic approach, which starts from the assumption that trust is rationally based. The analysis in the previous section, though, suggests that the social and cultural approach is also needed as a complement. When evaluating the probability of someone else's behavior, social and cultural factors cannot be ignored, mainly in the broader context – the delegation setting mentioned before – which does not refer to one specific decision situation. But we want to show here that, while the individualistic-rationalistic approach is incomplete to deal with the problem of trust, some of the characteristics of trust that are purported to be socially and culturally based are often a social reflection of the rationalistic approach.

Several researchers (e.g., Mayer et al., 1995) have emphasized the propensity to trust as one of the important characteristics that conditions actual trust, either suggesting that each person's personal experience is at the base of this propensity, or else pointing out that different cultural backgrounds differ in their propensity to trust (Hofstede, 1980).

As we have seen, Fischoff et al. (1988, pp. 399) argued that habits and traditions can be seen as trial-and-error learning summarized in rules and homilies, and that they provide stereotypic, readily justifiable responses to questions of values. This is particularly relevant in our context, because it means that the bases of trust can be considered to be instrumental from the beginning. Intuitions, rules and traditions provide an initial a priori probability *p* that trust will be honored by the other party, and any subsequent interactions of any kind (verbal or non-verbal communication, real actions between the two individuals, etc.) may modify that probability. Of course, for specific individuals this accumulated learning is transmitted only as a social habit or tradition.

Many of the individualistic-rationalistic models justify social beliefs and attitudes towards trust. Thus, Neilson (1999) develops a model where two agents interact repeatedly in a prisoner's dilemma, and shows that an agent A is willing to do costly favors for another agent B if A expects to receive favors in return in the future. His approach is quite clearly individualistic-rationalistic; of course, creating the social climate where one expects reciprocity in doing favors makes it easier for cooperation to exist. Along similar lines, Spagnolo (1999) shows how workers have an incentive to cooperate if there is a large enough probability that the other party will

cooperate. Valley, Moag and Bazerman (1998) show how, with asymmetric information, the communication medium (which is obviously a social creation) affects the distribution of outcomes, reflecting different degrees of truth-telling and trust across the media. Tullock (1999) changes the usual conditions of experiments using the prisoner's dilemma (i.e., he does not pre-select contestants, he does not prevent them from communicating and he does not change partners in the middle) and gets a very high degree of cooperation, in contrast with what happens under the usual conditions.

Thus, many unconscious habits may have their origins in rational attitudes. Dasgupta provides a good description of many of these factors from the rational perspective: "We form an opinion on the basis of his background, the opportunities he has faced, the courses of action he has taken, and so forth. Our opinion is thus based partly on the theory we hold of the effect of culture, class membership, family line, and the like on a person's motivation (his disposition) and hence his behavior" (Dasgupta, 1988, pp. 54).

But notice that two kinds of elements enter into this description. Background, culture, class membership, family line and so on may be considered (paradoxically perhaps) elements of the individualistic-rationalistic approach. They are elements that may indicate the kind of person the hypothetical trustee is, and what specifically can be expected from such a person: first-hand experiences, reputation, track record and so on.

In contrast, "the theory we hold of the effect of...." is clearly a cultural creation of the social group to which the would-be trustor belongs. Obviously, though, those social influences do not exhaust the explanation of one individual's propensity to trust, which by necessity must include personal factors.

The line between the individualistic-rationalistic point of view and the influence that the social environment exerts upon individuals is difficult to draw. Probability p is partly determined by the social background of the individual and partly by the individual's direct experience. Social, cultural and relational aspects of trust, however, have a background of instrumentality behind them.

Tyler and Degoey (1996, pp. 339) analyze in some depth the reasons why the instrumental view might not be enough to explain the phenomenon of trust. They give three arguments. First, if trust is merely instrumental, "people will care about trustworthiness when they are dependent on the organization or vulnerable to harm." Instead, what they found is that trustworthiness is central when people have "a personal connection with the authorities or identify with the organization." While this claim may of course be true, a personal connection with the authorities, or identification with the organization requires some knowledge of the authorities' value system, on which that trust can be based, according to this paper's analysis, on an instrumental basis.

Second, in the instrumental model, one would expect that trust would be "linked to satisfaction with the authority's decisions"; whereas if it is relational, it should be "linked to judgments about the neutrality of authorities and the degree to which these authorities treat their subordinates with dignity and respect." Again, under a rational-instrumental approach it is perfectly possible to argue that the value system of the authorities is at the root of the subordinates' trust in them.

Finally, if trust is instrumental in character, "judgments about the competence of authorities should be more strongly linked to people's willingness to accept an authority's decision than judgments about the benevolence of authorities." The way the problem of trust has been analyzed in this paper, that claim would have to be denied, given that the willingness to accept an authority's decision is related to trust in that authority, and that trust might, in fact, be based partly on competence and partly on value systems, a particular case of which is benevolence.

In summary, and as we have stated already, some of the characteristics of trust that are often assumed to be socially and culturally based may also be a social reflection of a rationalistic attitude.

4. Recapitulation and conclusions

Trust is a complex subject. It is elusive, as we have recognized from the beginning, and has many facets. It can have very different meanings, which we have tried to explore analytically in this paper. Now it is time to recapitulate and see where everything stands.

This paper started from the premise that trust is meaningful only if there is bounded rationality. In the context of this paper, this essentially stresses the point that agents do not have full knowledge of their own preferences and values, which completely changes the meaning of trust and the reasons for its existence. Under bounded rationality, preferences are organized in value systems, but the decisions made may or may not be consistent with them. A person's willingness to act according to the higher values (typically few in number) is one possible reason to trust that the person will follow some course of action that may be in that person's own best interest, though it may not be the most attractive in terms of the immediate variables of both effort and results. "Trust," then, means the belief by the trustor that the trustee will make the decision according to his/her real value system, even if some immediate variables push him/her in the opposite direction.

The next conclusion is rather intuitive. If, besides being consistent with a value system, some of the trustee's higher values are non-selfish, and include, for example, truthfulness, friendship, social welfare, etc., they provide a better basis for trust, i.e., the trustor may believe that the trustee will not take advantage of his/her vulnerabilities. In contrast with the previous situation, where values were not necessarily non-selfish, here the concept of trust may go beyond specific situations and extend to a class of decisions. This is, therefore, the concept that provides a foundation for taking some risks in situations of decentralization of authority, giving power to the trustee for a certain type of decisions.

Finally, whatever the actual preferences and values are, the trustee's action depends on his/her capacity to actually put into practice what he/she thinks is good according to his/her own system of values; therefore, trusting someone means trusting his/her capacity to do precisely that.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Aristotle (2000). *Nicomachean Ethics* (W. D. Ross, Trans.): The Internet Classics Archive by Daniel C. Stevenson. https://doi.org/10.1017/CBO9780511802058

Arrow, K.J. (1974). *The limits of Organization*. New York: Norton and Company.

Barnard, Ch.I. (1938). *The Functions of the Executive*. Cambridge, Mass: Harvard University Press.

Bazerman, M.H., Tenbrunsel, A.E., & Wade-Benzoni, K.. (1998). Negotiating with yourself and losing. *Academy of Management Review*, 23(2), 225-241. https://doi.org/10.5465/amr.1998.533224

Benabou, R., & Tirole, J. (2002). *Willpower and Personal Rules*. CEPR Discussion Paper No. 3143, January.

Bidault, F., & Jarillo, J.C. (1997). Trust in economic transactions. In P. Y. G. F. Bidault & a. G. Marion (Eds.), *Trust, Firm and Society* (pp. 81-94). London: Macmillan Business.

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 49-72). Oxford: Basil Blackwell.

Dietz, G. (2011). Going back to the source: Why do people trust each other?. *Journal of Trust Research*, 1(2), 215-222. https://doi.org/10.1080/21515581.2011.603514

Dirks, K.T., & Ferrin, D.L. (2001). The Role of Trust in Organizational Settings. *Organization Science,* 12(4), 450-467. https://doi.org/10.1287/orsc.12.4.450.10640

Ferrin, D.L., & Dirks, K.T. (2003). The Use of Rewards to Increase and Decrease Trust: Mediating Processes and Differential Effects. *Organization Science*, 14(1), 18-31. https://doi.org/10.1287/orsc.14.1.18.12809

Ferrin, D.L., & Gillespie, N. (2010). Trust differences across national-societal cultures: Much to do or much ado about nothing? In M. N. K. Saunders, D. Skinner, G. Dietz, N. Gillespie & R.J. Lewicki (Eds.), *Organizational trust: A cultural perspective* (pp. 42-86). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511763106.003

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1988). Knowing what you want: measuring labile values. In D. E. Bell, H. Raiffa & A. Tversky (Eds.), *Decision-making. Descriptive, normative and prescriptive interactions* (pp. 398-421). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511598951.020

Frey, B.S. (1997). *Not Just For the Money: An Economic Theory of Personal Motivation*. Brookfield: Edward Elgar.

Gabarro, J. (1978). The development of trust, influence and expectations. In A. G. Athos & J. Gabarro (Eds.), *Interpersonal Behavior: Communication and Understanding in Relationships* (pp. 290-303). Englewood Cliffs, NJ: Prentice-Hall.

Gambetta, D. (1988). Can We Trust Trust? In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 212-237). Oxford: Basil Blackwell.

Hirschman, A.O. (1984). Three easy ways of complicating economic discourse. *American Economic Review,* 74(2), 89-96.

Hofstede, G. (1980). Motivation, Leadership and Organization: Do American Theories Apply Abroad?. *Organizational Dynamics*, 9(1), 42-63. https://doi.org/10.1016/0090-2616(80)90013-3

Hosmer, L.T. (1995). Trust: the connecting link between organizational theory and philosophical inquiries. *Academy of Management Review*, 20, 379-403. https://doi.org/10.5465/amr.1995.9507312923

James, H.S. (2002). The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior and Organization*, 47, 291-307. https://doi.org/10.1016/S0167-2681(01)00214-1

Jensen, M. (2000). Value maximization, stakeholder theory and the corporate objective function. In M. Beer & N. Nohria (Eds.), *Breaking the Code of Change* (pp. 35-57). Boston, Mass: Harvard University Press. https://doi.org/10.2139/ssrn.220671

Jensen, M. (2002). *Just Say "No" to Wall Street. Negotiation, Organization and Markets Unit*, Harvard Business School. Working Paper No. 02-01.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59, 285-300. https://doi.org/10.1086/296367

Korczynski, M. (2000). The Political Economy of trust. *Journal of Management Studies*, 37, 1-21. https://doi.org/10.1111/1467-6486.00170

Kramer, R.M., & Tyler, T.R. (1996). Whither Trust? In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations* (pp. 1-15). Thousand Oaks: Sage Publications.

Kreps, D.M. (1990). Corporate Culture. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on Positive Political Economy* (pp. 90-143). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511571657.006

Lawler, E. (1969). Job design and employee motivation. *Personnel Psychology*, 22, 426-434. https://doi.org/10.1111/j.1744-6570.1969.tb00343.x

Lindenberg, S. (2001). Intrinsic Motivation in a New Light. *Kyklos*, 54(273), 317-342. https://doi.org/10.1111/1467-6435.00156

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior & Human Decision Processes*, 65(2), 272-292. https://doi.org/10.1006/obhd.1996.0028

March, J. (1987). Ambiguity and Accounting: The Elusive Link Between Information and Decision-Making. *Accounting, Organizations & Society*, 12, 153-168. https://doi.org/10.1016/0361-3682(87)90004-3

Mayer, R., Davis, J., & Schoorman, F. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, 20(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

McEvily, B. (2011). Reorganizing the Boundaries of Trust: From Discrete Alternatives to Hybrid Forms. *Organization Science*, 22(5), 1266-1276. https://doi.org/10.1287/orsc.1110.0649

McEvily, B., Perrone, V, & Zaheer, A. (2003). Trust as an Organizing Principle, *Organization Science: INFORMS: Institute for Operations Research*.

Neilson, W.S. (1999). The economics of favors. *Journal of Economic Behavior and Organizations*, 39, 387-397. https://doi.org/10.1016/S0167-2681(99)00047-5

O'Donoghue, T., & Rabin, M. (1999). Doing it Now or Later. *American Economic Review*, 89(1), 103-124. https://doi.org/10.1257/aer.89.1.103

Osterloh, M., & Frey, B.S. (2003). *Corporate Governance for Crooks? The Case for Corporate Virtue*. Working paper ISSN 1424-0459. Institute for Empirical Research in Economics, University of Zurich. https://doi.org/10.2139/ssrn.430062

Rokeach, M. (1973). *The Nature of Human Values*. New York: The Free Press.

Rosanas, J.M. (2016). *Is trust different from risk? Trust and trustworthiness with unbounded rationality: An extension of the trust paradox. IESE* Working paper.

Rousseau, D.M., Sitkin, S.B., Burt, R.S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404. https://doi.org/10.5465/amr.1998.926617

Russell, B. (1959). *The Problems of Philosophy*. Oxford: Oxford University Press.

Ryan, R.M., & Deci, E.L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25, 54-67. https://doi.org/10.1006/ceps.1999.1020

Saleh, S., & Hyde, J. (1969). Intrinsic vs Extrinsic Orientation and Job Satisfaction. *Occupational Psychology*, 43, 47-53.

Schelling, T. (1978). Egonomics or the Art of Self-Management. *American Economic Review*, 68(2), 290-294.

Schelling, T. (1984). Self-Command in Practice, in Policy and in the Theory of Rational Choice. *American Economic Review*, 74(2), 1-11.

Schoorman, F.D., Mayer, R.C., & Davis, J.H. (2007). An Integrative Model of Organizational Trust: Past, Present and Future. *Academy of Management Review,* 32(2), 344-354. https://doi.org/10.5465/amr.2007.24348410

Selten, R. (1999). *What is bounded rationality?* Discussion Paper B-454 Rheinische Friedrich-Wilhelms Universität, Bonn.

Senge, P. (2000). The Puzzles and Paradoxes of How Living Companies Create Wealth. Why Single-Valued Objective Functions Are Not Quite Enough. In M. Beer & N. Nohria (Eds.), *Breaking the Code of Change* (pp. 59-81). Boston, Mass.: Harvard University Press.

Simon, H.A. (1957). *Administrative Behavior* (Second Edition ed.). New York: The Free Press.

Simon, H.A. (1983). *Reason in Human Affairs*. Oxford: Basil Blackwell.

Simon, H.A. (1987). Making Management Decisions: the Role of Intuition and Emotion. *Academy of Management Executive*, 1, 57-64.

Simon, H.A. (1997). *Administrative Behavior* (fourth edition ed.). New York: The Free Press.

Spagnolo, G. (1999). Social relations and cooperation in organizations. *Journal of Economic Behavior and Organizations*, 38, 1-25. https://doi.org/10.1016/S0167-2681(98)00119-X

Tomlinson, E.C., & Mayer, R.C. (2009). The Role of Causal Attribution Dimensions in Trust Repair. *Academy of Management Review*, 34(1), 85-104. https://doi.org/10.5465/amr.2009.35713291

Tullock, G. (1999). Non-prisoner's dilemma. *Journal of Economic Behavior and Organizations*, 39, 455-458.

Tyler, T.R. , & Degoey, P. (1996). Trust in organizational authorities: the influence of motive attributions on willingness to accept decisions. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations* (pp. 331-356). Thousand Oaks: Sage Publications. https://doi.org/10.4135/9781452243610.n16

Valley, K.L., Moag, J., & Bazerman, M.H. (1998). A matter of trust: Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organizations*, 34, 211-238.

Webster's Seventh New Collegiate Dictionary. Webster (Eds) (1972). Springfield, Mass.: G&C Merriam Co.

Williams, M. (2007). Building Genuine Trust Through Interpersonal Emotion Management: A Threat Regulation Model of Trust and Collaboration Across Boundaries. *Academy of Management Review*, 32(2), 595-621. https://doi.org/10.5465/amr.2007.24351867

Williamson, O.E. (1993). Calculativeness, trust and economic organization. *Journal of Law and Economics*, 34, 453-502. https://doi.org/10.1086/467284

Williamson, O.E. (1985). *The Economic Institutions of Capitalism*. New York: The Free Press.

Zand, D.E. (1972). Trust and Managerial Problem Solving. *Administrative Science Quarterly*, 17, 229-239. https://doi.org/10.2307/2393957

Appendix: Possible list of psychological states associated with not knowing what you want